Title

Optimized gene expression from bacterial chromosome by high-throughput integration and screening

Short title: Optimizing expression by Tn5 integration

Authors

Tatyana E. Saleski[1], Meng Ting Chung[2], David N. Carruthers[1], Azzaya Khasbaatar[1], Katsuo Kurabayashi[2,3], Xiaoxia Nina Lin[1,4]*

Affiliations

[1]Department of Chemical Engineering, University of Michigan, Ann Arbor, MI, USA.
[2]Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI, USA.
[3]Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA.
[4]Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI, USA.

*Corresponding author. Email: ninalin@umich.edu.

Abstract

Chromosomal integration of recombinant genes is desirable compared to expression from plasmids due to increased stability, reduced cell-to-cell variability, and the elimination of antibiotics for plasmid maintenance. Here, we present a new approach for tuning pathway gene expression levels via random integrations and high-throughput screening. We demonstrate multiplexed gene integration and expression-level optimization for isobutanol production in *Escherichia coli*. The integrated strains could, with significantly lower expression levels than plasmid-based expression, produce high titers (10.0 ± 0.9 g/L isobutanol in 48 h) and yields (69 % of the theoretical maximum). Close examination of pathway expression in the top-performing, as well as other isolates, reveals the complexity of cellular metabolism and regulation, underscoring the need for precise optimization while integrating pathway genes into the chromosome. This new method for multiplexed pathway gene integration and expression optimization could be readily extended to a wide range of pathways and chassis to create robust and efficient production strains.

MAIN TEXT

Introduction

Microbial biosynthesis is a sustainable, high-specificity approach to achieving chemical conversions with the potential to produce a vast assortment of pharmaceutical, fuel, and commodity marginchemicals. Development of a high-performing production strain for a desired molecule generally requires tuning the expression levels of native and/or heterologous genes in hosts such as *Escherichia coli*[1, 2]. While extrachromosomal multicopy plasmids offer a convenient method for rapidly prototyping different expression levels of genes, and the majority of metabolic engineering efforts utilize them, they have several challenging attributes. First and foremost, plasmids are unstable, which has been extensively documented and studied due to the importance of plasmids in biotechnology[3-9]. Cells with plasmids can suffer from both structural instability, in which the plasmid is still carried but mutations inactivate the gene of interest, as well as segregational instability, in which some cells no longer carry the plasmid[10, 11]. Plasmids can also multimerize, which decreases their stability[12, 13]. Second, plasmid copy number can vary between cells within stable populations, and the degree of variation is often not well characterized even for widely used plasmids[14]. Münch *et al.* presented a striking example of cellular heterogeneity during plasmid-based production of recombinant protein in *Bacillus megaterium*. The authors found that, under the condition of strong selective pressure, 30% of the cell population was in a low-producing state[15]. This heterogeneity was found to be a result of asymmetric plasmid distribution. Moreover, plasmids require a selective pressure, which is typically an antibiotic molecule. The addition of antibiotics raises the process costs and puts additional stress on the cells. It is also not always effective. Kanamycin, for example, can become an ineffective selection agent at high phosphate concentrations[16]. Furthermore, due to the rise of multidrug-resistant pathogens, there is strong motivation to reduce antibiotic usage[17].

63 Chromosomal integration and expression of genes is an effective way for circumventing the
64 issues of plasmids described above. This alternative approach leads to more stable and robust
65 production strains, as well as reduce leakiness of inducible expression systems, which are highly
66 desirable traits for large-scale and long-term production processes[18, 19]. Mairhofer *et al.* compared a
67 plasmid-based to a chromosomally integrated expression system and found a much more significant
68 stress response in the plasmid-carrying strain. They identified massive overtranscription of the
69 plasmid-based gene of interest, leading to diversion of ribosomes and other cellular resources, as
70 the source of the metabolic burden[20].

71 Significant research efforts have been devoted to developing techniques for synthetic
72 integrations[21-23]. The two tools most widely used in prokaryotes for targeted integration of constructs
73 are homologous recombination by expression of the RecET or λ-Red proteins[24-26]
74 and site-specific recombination[27]. Homologous recombination enables rapid integration into a
75 desired locus but is only suitable for constructs up to several kilobases in size and that do not carry
76 significant homology to another part of the genome. Site-specific recombination is less restrictive
77 of the input construct but requires a bacterial chromosomal attachment site (*attB*) in the recipient
78 genome, which can place limitations on the location of integration. This site restrictiveness can be
79 circumvented by first inserting the attachment site into the target locus[28]. Yet, despite the recognized
80 desirability of chromosomal integration and the genetic tools available for targeted integrations,
81 achieving suitable gene expression levels from the chromosome for optimal production remains far
82 more challenging than manipulation of gene expression via plasmids, even in well-studied
83 organisms such as *E. coli*[22].

84 Chromosomal gene expression levels are dependent on multiple determinants. Similar to
85 plasmid-based gene expression, elements of the genetic construct, including promoters, ribosome-
86 binding sites (RBSs), and enhancers or activators, can be exploited to optimize heterologous gene
87 expression. Yet, single-copy expression from the chromosome is generally weaker than that of the

88 same construct from a multi-copy plasmid. Subsequently, there have been efforts in increasing the
89 copy number on the chromosome to achieve high-level gene expression[18, 29].

90 Another determinant affecting the expression level from the chromosome strongly is the
91 location of gene integration. It has long been known that during exponential growth, genes closer to
92 the origin of replication experience higher gene dosage and, therefore, higher expression levels[30, 31].
93 More recently it has been shown that other factors of genomic position, such as level of DNA
94 compaction or proximity to active genes, can also affect gene expression levels[32, 33]. In examining
95 transcription levels of a reporter gene at various sites in the *E. coli* genome, Bryant *et al.*[32] and
96 Scholz *et al.*[33] observed significant differences, up to ~300-fold, in expression across the *E. coli*
97 genome, excluding gene dosage effects. Intriguingly, on the other hand, it has been suggested that
98 in nature, microorganisms may use transposon-mediated transfer of catabolic genes to different
99 genomic locations as a means of fine-tuning their expression levels in order to adapt to new
100 environmental conditions[34]. These findings suggest that modulation of gene integration position can
101 serve as a useful tool for regulating expression levels. For instance, Loeschcke *et al.* have developed
102 an approach for transfer and expression (TREX) of biosynthetic pathways in bacteria utilizing Tn5
103 transposase to integrate gene clusters into random sites of the chromosomes and, based on resulting
104 colony color intensities, observed a range of production levels for pigmented secondary
105 metabolites[35].

106 Synthetic biology applications targeting the production of a small molecule often require
107 delicately balanced cellular resource allocation between production and growth, as well as across
108 different genes of the biosynthetic pathway. It remains elusive to predict the optimal gene expression
109 levels for most pathways due to the highly entangled nature of cellular metabolic and regulatory
110 networks. Additionally, for chromosomal gene expression, it is not fully understood yet how the
111 expression of different regions of the chromosome is affected by changes in environmental
112 conditions and whether/how an inserted gene construct will affect the native expression level of the

113 insertion region. It is, therefore, very difficult to rationally decide where to integrate a gene
114 construct. Here, we present a new metabolic engineering strategy for optimizing a pathway's
115 performance by tuning the gene expression through position-dependent expression variation (Fig.
116 1). In this method, we utilize the Tn5 transposase to randomly integrate pathway genes in the *E. coli*
117 genome in a multiplexed fashion. We subsequently screen the libraries using syntrophic co-culture
118 amplification of production (SnoCAP), which converts the production phenotype into a readily
119 screenable growth phenotype, as described in our previous work[36]. We demonstrate the approach
120 using a pathway for the production of isobutanol, a promising drop-in biofuel[37].

121 Efficient isobutanol production in *E. coli* has been achieved by diverting the branched-chain
122 amino acid pathway intermediate 2-ketoisovalerate (2-KIV) for conversion into isobutanol[38] (Fig.
123 2A). It has been demonstrated that by overexpressing AlsS from *Bacillus subtilis*, IlvC and IlvD
124 from *E. coli* (for the conversion of pyruvate to 2-KIV), and KivD and AdhA from *Lactococcus lactis*
125 (for the conversion of 2-KIV to isobutanol), production can reach up to 84% of the theoretical yield
126 under aerobic batch conditions[38] and titers can reach up to 50.8 g/L under fed-batch conditions[39].
127 Resolution of a cofactor imbalance has enabled strains to reach 100% of theoretical yield under
128 anaerobic conditions[40], and a recently developed strategy for isobutanol-linked growth selection
129 under anaerobic conditions enabled optimization of pathway expression levels for achieving optimal
130 production rates[41].

131 While much work with this pathway has relied on plasmid-based expression of pathway
132 genes, there is also interest in developing chromosomally integrated strains. Akita et al. have
133 demonstrated the integration of the five pathway genes into the *E. coli* genome, resulting in a titer
134 of 6.8 g/L and a yield of 55%, without the addition of costly antibiotics or inducers[42]. Bassalo et al.
135 developed a new CRISPR-based method for high-efficiency, single-step chromosomal

136 In the present work, we demonstrate that the generation of libraries with isobutanol pathway
137 genes integrated into diverse chromosomal positions combined with high-throughput production

138 phenotype screening is an effective means of generating high-performance chromosomally
139 integrated strains. *E. coli* strains resulted from this study show, to our knowledge, the highest titers
140 and yields yet reported with the isobutanol pathway genes expressed exclusively from the
141 chromosome.

142 **Results**

143

144 *Construction and screening of Tn5 integration libraries*

145 We began by constructing an integration library (Library A) in which a construct containing
146 the *alsS* gene, under the PLlacO1 promoter, and a kanamycin-resistance gene, was integrated by
147 Tn5 transposase into the genome of JCL260 Δ*lysA* in random locations. This integration was
148 accomplished by PCR amplification of the *alsS* and *kan* genes with primers that add the 19 bp
149 mosaic end sequence recognized by Tn5, followed by reaction with Tn5, and electroporation of cells
150 with the resulting transposome. The approximate library size was determined to be ~$3.4 \pm 1.1 \times 10^4$
151 by plating a small quantity of the cells after a short initial recovery on kanamycin-containing plates.
152 The remainder of the library was subjected to a liquid kanamycin selection and subsequently
153 screened by SnoCAP. In the SnoCAP approach, the library is co-encapsulated in water-in-oil
154 microdroplets with a fluorescent sensor strain that is auxotrophic for the target molecule of interest.
155 The library itself is also auxotrophic for a second, orthogonal molecule supplied by the sensor strain.
156 This cross-feeding configuration enables conversion of the production phenotype into a fluorescent
157 output, amplified by co-culture growth[36]. To screen for overproducers of 2-ketoisovalerate, an
158 intermediate in the isobutanol pathway (Fig. 2A), we use a sensor strain with a deletion of the *ilvD*
159 gene (encoding a dihydroxy-acid dehydratase), which prevents it from converting 2,3-dihydroxy-
160 isovalerate into 2-ketoisovalerate. This strain can only grow when it is provided an exogenous
161 source of 2-ketoisovalerate and/or branched-chain amino acids. We demonstrated in our previous
162 work that this sensor can be used to identify strains that overproduce isobutanol once the genes
163 responsible for converting 2-ketoisovalerate to isobutanol (*kivD* and *adhA*) are added[36].

164 The cells are loaded at densities such that all droplets contain sensor cells, and most droplets
165 contain either zero or one secretor cell from the integration library. Thus, the co-growth phenotype
166 of individual library members may be assessed. The microdroplet compartmentalization format

167 enables high-throughput screening and isolation of droplets of interest by fluorescence-activated
168 droplet sorting (FADS)[43].

169 Fluorescence measurement of the droplets containing Library A members co-cultured with
170 fluorescent sensor strain showed that more droplets exhibited growth than when droplets contained
171 co-cultures with the parent strain, which lacks the *alsS* gene, as the secretor (Fig. 2Bi,ii). We sorted
172 the top 0.3% of droplets with the highest fluorescence signal and plated the pool of droplets on plates
173 with tetracycline to select for the secretor cells and against the sensor cells (the parent strain of the
174 secretor strain libraries contains a tetracycline resistance gene in the F plasmid, whereas the sensor
175 strain is tetracycline sensitive). We assessed the isobutanol production of colonies obtained from
176 the screening and colonies from the unscreened library, after transformation with pSA65. Of 12
177 colonies chosen at random from the unscreened library, 10 produced less than 0.1 g/L after 48 h of
178 fermentation (similar to the parent strain), one produced 0.4 g/L and the other produced 2.1 g/L (Fig.
179 S1). Of 10 random colonies tested from the sorted pool, all produced more than 2.5 g/L (Fig. S1),
180 which demonstrates the effectiveness of the screening method.

181 Because transposome integration efficiency is expected to decrease with increasing construct
182 length, we chose to use the *ilvCD* genes as a selection marker by utilizing their essentiality in
183 minimal medium, rather than an antibiotic. We deleted *ilvD* from JCL260 Δ*lysA* and integrated
184 constructs containing either both genes *ilvC* and *ilvD* behind the PLlacO1 promoter or only the *ilvD*
185 gene behind the same promoter. We performed the selection in minimal M9 medium supplemented
186 with IPTG and lysine, with a small amount plated on solid medium for library size estimation and
187 the rest of the library undergoing liquid selection. These integrations produced far smaller library
188 sizes than the *alsS* integration construct (~10^2 integrants per electroporation vs. ~10^4 for *alsS*). We
189 determined that this was due to substantially lower survival rates in the minimal medium selection
190 compared to kanamycin selection in LB medium. For example, when JCL260 Δ*lysA* Δ*ilvD* was
191 transformed with pSA69, the number of colonies recovered on LB plates with kanamycin was two

192 orders of magnitude higher than the number on M9 plates with lysine and IPTG. Further
193 optimization of the recovery method, such as supplementation of additional nutrients or seeding of
194 small quantities of the branched-chain amino acids may help improve this lower-than-expected
195 recovery rate and increase the diversity of the library. It should also be noted that the survival rate
196 on the solid medium may be different from that in the liquid medium, which could contribute to
197 inaccuracies in estimated library size. Deep sequencing of the library could be used to quantify the
198 library size more accurately. We repeated the *ilvD* and *ilvCD* library generation several times and
199 pooled the integrants, resulting in libraries of each of size ~500-1,000 (estimated from agar plating).
200 We then integrated the *alsS-kan* transposome into these libraries, which led to ~$1.1 \pm 0.5 \times 10^4$
201 integrants for the *ilvD/alsS* library (Library B) and $1.3 \pm 0.3 \times 10^4$ for the *ilvCD/alsS* library (Library
202 C). Therefore, the library should consist of about 10 different *alsS* integration locations for each
203 *ilvD* or *ilvCD* integration library member.

204 When assessed by the SnoCAP method (library members co-encapsulated in microdroplets
205 with sensor strain, incubated for co-growth, and the fluorescence of the droplets measured), the *ilvD*
206 and *ilvCD* integration libraries displayed no measurable co-growth. The *ilvD/alsS* and *ilvCD/alsS*
207 libraries (Libraries B and C, respectively) on the other hand, showed substantially more fluorescence
208 compared to the library with only *alsS* integrated (Library A) (Fig. 2B).

209 We sorted the droplets containing co-cultures of these libraries and recovered colonies from
210 the pools of droplets, and then transformed the isolates with plasmid pSA65 (encoding KivD and
211 AdhA) and assessed isobutanol production. The best isolates were found to be from Library C. In
212 particular, an isolate termed C7 reached a production level closely approaching that of the strain
213 expressing the pathway genes from a plasmid (Fig. 3A). With the goal of understanding the
214 relationship between production phenotype and genome-position-dependent gene expression, we
215 decided to examine isolates displaying a spectrum of production levels and hence also isolated

216 several library members with low- and intermediate-production levels (Fig. 3A) by screening the
217 random libraries via the microplate format of SnoCAP.
218
219 *Genomic characterization of select library isolates*
220
221 We then determined the insertion site locations in the selected isolates by transposon
222 footprinting, which also allowed us to eliminate duplicates of the same genotype and resulted in a
223 panel of unique isolates reported in this work. The integration locations were found to spread across
224 the whole genome (Fig. 3B, Table 1). Intriguingly, each of the Library B isolates, B1 and B2,
225 contains integrations into the F plasmid, a large (~100 kb) plasmid previously introduced to the base
226 strain to supply *lacIq* for increased levels of the lac repressor. The F plasmid is stably maintained
227 in *E. coli* due to an active partitioning system[44, 45], and it has been suggested that placing
228 heterologous genes on the F plasmid may provide improved stability over the use of other
229 plasmids[46, 47]. It is also interesting that in isolate B2, both integration constructs are found to be
230 located on the F plasmid.
231 Another noteworthy observation was that two of the disrupted genes, *ttdT* and *yjhI*, are genes
232 in which we had found mutations during previous work in which we chemically mutagenized
233 JCL260 Δ*lysA*, screened for improved isobutanol productivity and resequenced the genome of an
234 improved isolate[36]. In that study, the strain with the best production level possessed a mutation in
235 the *ttdT* gene leading to an alanine to valine substitution at position 390 of the protein, and *yjhI*
236 contained a mutation leading to a premature stop codon after the first 187 amino acids (of 227).
237 In the best isolate, C7, the two integration constructs landed in *ttdT* and *yqiH,* two genes
238 that are quite close to each other, separated by only 18 kb (Fig. 4A). To understand the genetic
239 mechanism underlying the high-production phenotype of this isolate, we investigated whether the
240 gene disruptions, besides expression of the inserted genes, can contribute to increased production.
241 We knocked out the *ttdT* and *yqiH* genes individually in a base strain JCL260 Δ*lysA*.

242

243 Into each deletion strain, we then integrated *alsS* in the *yghX* site, which we have previously seen
244 leads to an intermediate production level (and therefore should allow us to see changes in the
245 production that may not be visible in a very low- or high-producing background strain). After
246 transformation of the strains with pSA65, we tested isobutanol production and found that the *ttdT*
247 deletion strain produces more isobutanol to a statistically significant degree compared to the strain
248 possessing wildtype *ttdT*, whereas *yqiH* deletion does not have a statistically significant effect (Fig.
249 4B).

250
251 *Characterization of pathway gene transcript levels in select library isolates*

252 We next analyzed gene expression in various isolates of different production levels. We
253 harvested the RNA during exponential growth (3 h after induction of the pathway genes with
254 IPTG), reverse transcribed it and analyzed gene expression of *alsS*, *ilvC*, and *ilvD* by quantitative
255 real-time PCR (qPCR) (Fig. 4D), while meanwhile continuing to grow the cultures and measuring
256 their isobutanol production 48 h following induction (Fig. 4C). It was found that the pathway
257 gene expression in each of the integrated strains was, as expected, far lower than that in the strain
258 expressing *alsS/ilvCD* from the pSA69 plasmid (Fig. 4D, note the log-scale of the Y-axis). Yet,
259 with significantly lower expression levels of all three genes, the top isolate C7 showed a
260 production level very close to that of the plasmid strain.

261 With the gene transcript profiles of these isolates, we sought to elucidate the relationship
262 between gene expression and production phenotype. Some interesting observations were made.
263 For instance, predictably, isolates from Library A have lower expression of *ilvD* than isolates
264 from Libraries B and C. However, *ilvC* can be seen to be upregulated even in certain strains that
265 do not have *ilvC* insertion (A3, A4, B1). These strains have high expression levels of *alsS*, so it
266 is likely that accumulation of acetolactate, the product of AlsS, is inducing higher transcription
267 levels of *ilvC* in these strains, as has been previously observed[48]. We also noted that there was a

268 similar trend of relative expression levels within the pathway (*ilvC* highest followed by *alsS* and
269 then *ilvD*) in the medium- and high-production isolates (A3, A4, B1, C5, C6, C7), which was
270 also seen in the plasmid strain. Most strikingly, however, there was a lack of a clear relationship
271 between gene expression levels and production. In particular, despite the fact that the top four
272 analyzed isolates (B1, C5, C6, C7) all showed a much higher expression of *ilvD* than the next
273 two, substantially lower-producing, isolates (A3, A4), it is not clear what causes C7 to perform
274 significantly better in production than the other three isolates or why differences in gene
275 expression levels across these three do not lead to differences in production. Examination of the
276 low-production isolates also provided some insights. Low expression, especially of *alsS*,
277 appeared to have caused low production in A1 and A2. Intriguingly, another isolate C1 showed
278 much higher expression of the three pathway genes but did not exhibit increased production. We
279 suspect that this could be due to deleterious effects of disrupting the genes in the integration sites
280 or to poor pathway balancing. The above findings underscore the highly entangled and complex
281 nature of optimizing pathway gene expression for a specific production phenotype.
282
283 *Construction and characterization of strains with all isobutanol pathway genes integrated*
284 To generate strains in which the entire isobutanol pathway is expressed from the chromosome,
285 we integrated the *adhA* and *kivD* genes into isolate C7. Because in the work of Atsumi et al.[38],
286 these genes were expressed from a high-copy-number plasmid, we expected that high expression
287 levels would be needed to achieve suitable production levels. We utilized the chemically
288 inducible chromosomal evolution (CIChE) method of Tyo et al.[18] for generating multiple copies
289 of a gene in a single integration site. In this approach, the genes of interest are placed alongside
290 an antibiotic marker between 1 kb homology regions that enables RecA-mediated recombination
291 to increase the copy number in response to increases in antibiotic concentration (Fig. 5A). If
292 desired, the copy number can be stabilized by deletion of the *recA* gene. We integrated the CIChE

293 construct containing *kivD, adhA,* and the chloramphenicol resistance gene *cat* into the *aslB* site,
294 a high-expression site[32]. Before integrating this construct into C7, we also integrated it into a
295 clean background strain, passaged it into higher chloramphenicol concentrations, deleted the
296 *recA* gene to stabilize the copy number, and measured the construct copy number by qPCR of the
297 *cat* gene (Fig. S2). These measurements demonstrated that the copy number can be increased up
298 to about 60 copies by this method. We tested the production of the C7 strain with the CIChE
299 *kivD-adhA* integration (C7 CIChE) in 50 μg/mL chloramphenicol and found that it produced
300 more isobutanol than a strain with just a single copy of *kivD* and *adhA* integrated into the same
301 site. Further increasing the chloramphenicol concentration did not lead to an increase in
302 production level. The fully integrated strain C7 CIChE performed very similarly to C7 pSA65,
303 generating isobutanol titers and yields lower than but close to those of the double plasmid strain
304 (Fig. 5B,C). Noticeably, on the other hand, this strain reached higher cell densities than the
305 plasmid-bearing strains (Fig. 5D).

306 We also examined the *kivD* and *adhA* transcript levels in the fully integrated strain
307 compared to the plasmid-bearing strains. As expected, they are lower in C7 CIChE. Quite
308 surprisingly, however, we found them to be dramatically lower (more than two orders of
309 magnitude) in the integrated strain compared to those in both strains expressing *kivD*/*adhA* from
310 plasmid pSA65 (Fig. 5E). This indicates that the expression levels rendered by the high-copy
311 number plasmid are unnecessarily high. This result further demonstrates that it is feasible to
312 achieve high production through significantly lower pathway gene expression from the
313 chromosome.

314

315 **Discussion**

316

317 Synthetic biology is making strides toward solutions to some of society's most pressing
318 problems. Managing the issue of stability and cell-to-cell variability, however, still present major
319 hurdles to overcome[49]. Moving gene expression from plasmids to the genome can improve stability

320 and reduce cell-to-cell variability. Yet it remains challenging to predict the expression levels that
321 will be optimal for production and achievable by genomic integration into a particular site.
322 Moreover, the interplay between gene integration and surrounding regions on the chromosome, such
323 as gene disruption and expression level perturbation, could cause complications and additional
324 difficulty in optimizing pathway expression.

325 The new approach of screening of random integration libraries developed in the present
326 study circumvents the challenges of rational selection of integration locations. The transposon
327 integration method is versatile and applicable to almost any gene of interest. Since the integration
328 is not homology-based (unlike lambda Red recombinase-based integration, for example) it can even
329 be used on genes that are already natively present in the genome at suboptimal expression levels,
330 such as *ilvC* in the work presented here. The high-throughput screening enables large libraries to be
331 assessed, making it feasible to multiplex integrations of different pathway genes.

332 The application of the approach to the isobutanol pathway demonstrates its effectiveness and
333 has generated a strain with, to our knowledge, the highest production titers yet reported in the
334 literature for an *E. coli* strain with the isobutanol pathway chromosomally integrated. This
335 chromosomally integrated strain, which achieves production performance close to that of a plasmid-
336 based strain with substantially higher pathway gene expression levels, also provides a striking
337 example of how plasmid-based expression can be unnecessarily high. We expect that the lower
338 expression levels in the integrants may be beneficial under more stressful conditions, alleviating
339 unnecessary demands on the cell. Even under the relatively optimal fermentation conditions used
340 here, the integrated strains reach higher cell densities than the plasmid-bearing strains,
341 demonstrating the reduced burden on these strains. Furthermore, we found this approach can take
342 advantage of beneficial gene deletions that may not be obvious choices for targeted deletion to yield
343 optimal production.

Our results also open intriguing questions for future investigation. For example, it will be interesting to learn what factors can lead a strain with relatively high expression levels of all pathway genes, such as C1, to have such a substantially lower production level than other strains. Also intriguing is the fact that, in the highest-producing isolate we identified, C7, as well as in another high-producing isolate, B2, the gene integrations are located quite close to each other. Whether this is due to that region having optimal expression levels, or the proximity of the constructs provides some benefit, or it is merely by chance is not clear. Deep-sequencing pools of droplets collected from different degrees of screening stringency could provide further insights into whether proximity leads to improvements in production. Additionally, studies with targeted integrations could be designed to address this question.

357
358 **Materials and Methods**
359
360 *Strains and plasmids*

361 Strains and plasmids used are listed in Table S1. JCL260, pSA65, and pSA69[38, 50] were provided by
362 James Liao, UCLA. Keio strains were obtained from the *E. coli* genetic stock center (CGSC,
363 http://cgsc2.biology.yale.edu/). Plasmid pSAS31 was constructed by Scott Scholz[33].

364

365 *Media*

366 M9IPG, consisting of M9 salts (47.8 mM Na_2HPO_4, 22.0 mM KH_2PO_4, 8.55 mM NaCl, 9.35 mM
367 NH_4Cl, 1 mM $MgSO_4$, 0.3 mM $CaCl_2$), micronutrients (2.91 nM $(NH_4)_2MoO_4$, 401.1 nM H_3BO_3,
368 30.3 nM $CoCl_2$, 9.61 nM $CuSO_4$, 51.4 nM $MnCl_2$, 6.1 nM $ZnSO_4$, 0.01 mM $FeSO_4$), thiamine HCl
369 (3.32 μM) and dextrose (D-glucose) at the stated concentrations, was used for all culturing
370 experiments. For SnoCAP screening, in both microdroplets and microplates, the glucose
371 concentration was 20 g/L, and the medium was supplemented with 3 mM L-isoleucine and 50
372 μg/mL kanamycin. For isobutanol production monocultures, the medium contained 36 g/L glucose,
373 5 g/L yeast extract, and no antibiotics were added. Precultures for both screening and production
374 cultures were carried out in LB Lennox with antibiotics appropriate to the strain. When used,
375 antibiotics were supplied at the following concentrations: ampicillin, 100 μg/mL; kanamycin, 50
376 μg/mL; tetracycline, 10 μg/mL; chloramphenicol, 50 μg/mL.

377

378 *Transposon integration library construction*

379 All primers and oligonucleotides were ordered from Integrated DNA Technologies, Inc. (IDT) and
380 are listed in Table S2. Each of the integration constructs was amplified by PCR from the pSA69
381 plasmid, adding the Tn5 mosaic ends. Primers phosph_transp_kan_for and phosph_transp_alsS_rev
382 were used to amplify *kan*-PLlacO1-*alsS* (for Libraries A, B and C).
383 phosph_transp_PLlacO1_ilvC_for and phosph_transp_ilvD_rev were used to amplify *ilvCD* and

384 introduce a PLlacO1 promoter in front of *ilvC* (for Library C). phosph_transp_PLlacO1_ilvD_for
385 and phosph_transp_ilvD_rev were used to amplify *ilvD* and introduce a PLlacO1 promoter in front
386 of *ilvD* (for Library B). The linear PCR product was then digested with both DpnI and SpeI enzymes
387 (NEB) to digest the template plasmid, phosphorylated with T4 polynucleotide kinase (NEB),
388 cleaned with a PCR clean-up kit (Qiagen) and eluted in TE buffer. The DNA was then reacted with
389 EZ-Tn5 transposase (Lucigen) according to the manufacturer's instructions. The resulting
390 transposome was then electroporated into the appropriate strain. The cells were recovered for 1.5 hr
391 with 1 mL SOC medium. Then 50 µL was used for dilution and plating on LB with kanamycin
392 plates to assess library size, and the remaining cells were grown to saturation in 100 mL selective
393 medium (either LB with 50 µg/mL kanamycin for *alsS* integration or M9IPG with 0.1 mM IPTG
394 and 3 mM lysine for *ilvCD* or *ilvD* integration). The cells were then frozen in 1 mL aliquots
395 (resuspended in fresh LB with 25% glycerol) and later thawed, washed, and grown in LB with 50
396 µg/mL kanamycin to prepare them for screening.
397
398 *SnoCAP screening*
399 SnoCAP screening for 2-KIV production, in droplet and microplate formats, was employed as
400 previously described[36]. Stationary phase cultures in LB were used as the inocula for both formats.
401 For the microplate format, K12 Δ*ilvD* was used as the sensor strain. M9IPG with 20 g/L glucose, 3
402 mM isoleucine, and 50 µg/mL kanamycin was used as the medium for all screening co-cultures
403 except for those with the *ilvD* and *ilvCD* libraries (without *alsS* integration), in which the kanamycin
404 was omitted. For the droplet format, K12 Δ*ilvD* pSAS31, which expresses mNeonGreen, was used
405 as the sensor strain. For the droplet sorting assay, the droplet collection device was soaked in a
406 mixture of HFE-7500 oil and water for several days before use in order to improve droplet stability
407 after collection.
408

409 *Isobutanol fermentations and analysis*

410 Isobutanol cultures were grown similarly to the method previously described[36]. Overnight cultures
411 in LB with antibiotics were diluted 1:100 v/v into 10 mL of M9IPG with 36 g/L glucose, 5 g/L yeast
412 extract, and no antibiotics, in a 125-mL baffled, unvented polypropylene flasks. Cells were grown
413 for 2.5 to 3 h at 37 °C, 250 rpm, followed by induction of the pathway genes with 0.1 mM IPTG.
414 Flasks were then sealed with parafilm and incubated at 30 ^{o}C, 250 rpm. Samples were taken daily
415 to measure cell density and glucose and isobutanol concentrations. Cell density was assessed by
416 measuring the OD_{600} of 200 μL of culture, diluted into the linear range, in a VersaMax microplate
417 reader (Molecular Devices). Glucose and isobutanol were measured by high-performance liquid
418 chromatography (HPLC), as previously described[37]. In the case of cultures in which RNA was
419 harvested, 0.5 mL of cells were harvested with RNAprotect Bacteria Reagent (Qiagen) 3-4 h after
420 induction. RNA was isolated using an RNA mini kit (Qiagen) according to the manufacturer's
421 instructions, including on-column DNAse I digestion. An additional digestion using Turbo DNA-
422 free Kit (Invitrogen) was performed after purification to eliminate any remaining genomic DNA.
423 The RNA was then reverse transcribed to cDNA using MultiScribe Reverse Transcriptase
424 (Invitrogen) and approximately 400 ng RNA per 20 μL reaction, with the following temperature
425 profile: 25 ^{o}C, 10 min; 37 ^{o}C, 120 min, 85 ^{o}C, 5 min. A no reverse transcriptase control was carried
426 out for each sample and checked with at least one set of primers during the real-time PCR analysis
427 to ensure no significant residual genomic DNA. qPCR was conducted on the cDNA using SYBR
428 Green qPCR Master Mix (Life Technologies) with 20 μL reactions. Reactions were run in 96-well
429 plates on a 7900HT Fast Real-Time PCR Machine (Applied Biosystems), courtesy of the University
430 of Michigan Advanced Genomics Core. The PCR program was as follows: 50 ^{o}C, 2 min; 95 ^{o}C, 10
431 min; 40 cycles of 95 ^{o}C for 15 s, 60 ^{o}C for 1 min. Primers used are listed in Table S2. Data were
432 analyzed using the $2^{-\Delta\Delta CT}$ method[51].

433

434 *kivD and adhA integration*

435
436 A CIChE construct consisting of *kivD, adhA*, and a chloramphenicol resistance gene, *cat*, flanked
437 by 1 kb homology regions, was previously generated and integrated into the *aslB* locus of the
438 NV3r1 strain by λ-Red recombineering[36]. Following the method of Tyo et al.[18], we passaged this
439 strain into increasing concentrations of chloramphenicol and then deleted the *recA* gene by P1
440 transduction from donor strain BW 26,547 Δ*recA*::*kan* Lambda *recA*+ (obtained from the Coli
441 Genetic Stock Center (CGSC)). Transductants were selected on LB plates with 50 μg/mL
442 kanamycin and the corresponding concentration of chloramphenicol. Genomic DNA was isolated
443 using DNEasy Blood & Tissue Kit (Qiagen), and the copy number was determined by qPCR using
444 the same primers as Tyo et al.[18], with one set for the *cat* gene and one for the *bioA* gene, which is
445 used as a single copy reference gene. Absolute copy number was determined by comparison to a
446 strain with a single copy of *cat* integrated into the *intC* locus (NV3r1 Δ*intC::yfp-cat*). To transfer
447 this integration into the C7 strain generated in this study, we prepared P1 lysates from NV3r1
448 (*recA*+) with the *kivD, adhA* CIChE construct after growing on 80 μg/mL chloramphenicol and
449 transduced it into the C7 strain, selecting on LB plates with 40 μg/mL chloramphenicol. Resulting
450 colonies were isolation streaked twice on LB agar with 40 μg/mL chloramphenicol and 0.8 mM
451 sodium citrate.
452
453
454

455
456 **References**
457

458 1. Woolston, B.M., Edgar, S. & Stephanopoulos, G. Metabolic Engineering: Past and Future.
459 *Annual Review of Chemical and Biomolecular Engineering* **4**, 259–288 (2013).
460 2. Nielsen, J. & Keasling, Jay D. Engineering cellular metabolism. *Cell* **164**, 1185-1197
461 (2016).
462 3. Popov, M., Petrov, S., Nacheva, G., Ivanov, I. & Reichl, U. Effects of a recombinant gene
463 expression on ColE1-like plasmid segregation in *Escherichia coli*. *BMC Biotechnology* **11**,
464 18 (2011).
465 4. Bentley, W.E. & Kompala, D.S. Plasmid instability in batch cultures of recombinant
466 bacteria. A laboratory experiment. *Chemical Engineering Education* **24**, 168-172 (1990).
467 5. Seo, J.H. & Bailey, J.E. A segregated model for plasmid content and product synthesis in
468 unstable binary fission recombinant organisms. *Biotechnology and Bioengineering* **27**,
469 156-166 (1985).
470 6. Siegel, R. & Ryu, D.D. Kinetic study of instability of recombinant plasmid pPLc23trpAl in
471 *E. coli* using two-stage continuous culture system. *Biotechnology and Bioengineering* **27**,
472 28-33 (1985).
473 7. Imanaka, T., Tsunekawa, H. & Aiba, S. Phenotypic stability of trp operon recombinant
474 plasmids in *Escherichia coli*. *Microbiology* **118**, 253-261 (1980).
475 8. Imanaka, T. & Aiba, S. A perspective on the application of genetic engineering: stability of
476 recombinant plasmid. *Annals of the New York Academy of Sciences* **369**, 1-14 (1981).
477 9. Ollis, D. Industrial fermentations with (unstable) recombinant cultures. *Philosophical*
478 *Transactions of the Royal Society of London. B, Biological Sciences* **297**, 617-629 (1982).
479 10. Friehs, K.P. "Plasmid copy number and plasmid stability" in New Trends and
480 Developments in Biochemical Engineering 47-82 (Springer, 2004).
481 11. Ensley, B.D. Stability of recombinant plasmids in industrial microorganisms. *Critical*
482 *Reviews in Biotechnology* **4**, 263-277 (1986).
483 12. Summers, D.K. & Sherratt, D.J. Multimerization of high copy number plasmids causes
484 instability: ColE 1 encodes a determinant essential for plasmid monomerization and
485 stability. *Cell* **36**, 1097-1103 (1984).
486 13. Field, C. & Summers, D. Multicopy plasmid stability: revisiting the dimer catastrophe.
487 *Journal of Theoretical Biology* **291**, 119-127 (2011).
488 14. Jahn, M., Vorpahl, C., Hübschmann, T., Harms, H. & Müller, S. Copy number variability
489 of expression plasmids determined by cell sorting and Droplet Digital PCR. *Microbial Cell*
490 *Factories* **15**, 211 (2016).
491 15. Münch, K.M. et al. Polar fixation of plasmids during recombinant protein production in
492 *Bacillus megaterium* results in population heterogeneity. *Applied and Environmental*
493 *Microbiology* **81**, 5976-5986 (2015).
494 16. Studier, F.W. Protein production by auto-induction in high-density shaking cultures.
495 *Protein Expression and Purification* **41**, 207-234 (2005).
496 17. Mignon, C., Sodoyer, R. & Werle, B. Antibiotic-free selection in biotherapeutics: now and
497 forever. *Pathogens* **4**, 157-181 (2015).
498 18. Tyo, K.E.J., Ajikumar, P.K. & Stephanopoulos, G. Stabilized gene duplication enables
499 long-term selection-free heterologous pathway expression. *Nature Biotechnology* **27**, 760-
500 765 (2009).
501 19. Striedner, G. et al. Plasmid-free T7-based *Escherichia coli* expression systems.
502 *Biotechnology and Bioengineering* **105**, 786-794 (2010).
503 20. Mairhofer, J., Scharl, T., Marisch, K., Cserjan-Puschmann, M. & Striedner, G.
504 Comparative transcription profiling and in-depth characterization of plasmid-based and

505 plasmid-free *Escherichia coli* expression systems under production conditions. *Applied*
506 *and Environmental Microbiology* **79**, 3802-3812 (2013).
507 21. Madyagol, M. et al. Gene replacement techniques for Escherichia coli genome
508 modification. *Folia Microbiologica* **56**, 253-263 (2011).
509 22. Ou, B., Garcia, C., Wang, Y., Zhang, W. & Zhu, G. Techniques for chromosomal
510 integration and expression optimization in Escherichia coli. *Biotechnology and*
511 *Bioengineering* **115** (2018).
512 23. Li, L., Liu, X., Wei, K., Lu, Y. & Jiang, W. Synthetic biology approaches for
513 chromosomal integration of genes and pathways in industrial microbial systems.
514 *Biotechnology Advances* **37**, 730-745 (2019).
515 24. Zhang, Y., Buchholz, F., Muyrers, J.P. & Stewart, A.F. A new logic for DNA engineering
516 using recombination in *Escherichia coli*. *Nature Genetics* **20**, 123-128 (1998).
517 25. Yu, D., Ellis, H.M., Lee, E.-C., Jenkins, N.A. & Copeland, N.G. An efficient
518 recombination system for chromosome engineering in *Escherichia coli*. *Proceedings of the*
519 *National Academy of Sciences* **97**, 5978-5983 (2000).
520 26. Datsenko, K.A. & Wanner, B.L. One-step inactivation of chromosomal genes in
521 Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of*
522 *Sciences* **97**, 6640-6645 (2000).
523 27. Grindley, N.D., Whiteson, K.L. & Rice, P.A. Mechanisms of site-specific recombination.
524 *Annual Review of Biochemistry* **75**, 567-605 (2006).
525 28. Minaeva, N.I. et al. Dual-In/Out strategy for genes integration into bacterial chromosome:
526 a novel approach to step-by-step construction of plasmid-less marker-less recombinant *E.*
527 *coli* strains with predesigned genome structure. *BMC Biotechnology* **8**, 63 (2008).
528 29. Yin, J. et al. Effects of chromosomal gene copy number and locations on
529 polyhydroxyalkanoate synthesis by *Escherichia coli* and *Halomonas* sp. *Applied*
530 *Microbiology and Biotechnology* **99**, 5523-5534 (2015).
531 30. Beckwith, J.R., Signer, E.R. & Epstein, W. in Cold Spring Harbor symposia on
532 quantitative biology, Vol. 31 393-401 (Cold Spring Harbor Laboratory Press, 1966).
533 31. Schmid, M.B. & Roth, J.R. Gene location affects expression level in *Salmonella*
534 *typhimurium*. *Journal of Bacteriology* **169**, 2872-2875 (1987).
535 32. Bryant, J.A., Sellars, L.E., Busby, S.J. & Lee, D.J. Chromosome position effects on gene
536 expression in *Escherichia coli* K-12. *Nucleic Acids Research* **42**, 11383-11392 (2014).
537 33. Scholz, S.A. et al. High-resolution mapping of the *Escherichia coli* chromosome reveals
538 positions of high and low transcription. *Cell Systems* **8**, 212-225 (2019).
539 34. Sousa, C., de Lorenzo, V. & Cebolla, A. Modulation of gene expression through
540 chromosomal positioning in *Escherichia coli*. *Microbiology* **143**, 2071-2078 (1997).
541 35. Loeschcke, A. et al. TREX: a universal tool for the transfer and expression of biosynthetic
542 pathways in bacteria. *ACS Synthetic Biology* **2**, 22-33 (2012).
543 36. Saleski, T.E. et al. Syntrophic co-culture amplification of production phenotype for high-
544 throughput screening of microbial strain libraries. *Metabolic Engineering* **54**, 232-243
545 (2019).
546 37. Chen, C.-T. & Liao, J.C. Frontiers in microbial 1-butanol and isobutanol production.
547 *FEMS Microbiology Letters* **363**, fnw020 (2016).
548 38. Atsumi, S., Hanai, T. & Liao, J.C. Non-fermentative pathways for synthesis of branched-
549 chain higher alcohols as biofuels. *Nature* **451**, 86 (2008).
550 39. Baez, A., Cho, K.-M. & Liao, J.C. High-flux isobutanol production using engineered
551 *Escherichia coli*: a bioreactor study with in situ product removal. *Applied Microbiology*
552 *and Biotechnology* **90**, 1681-1690 (2011).
553 40. Bastian, S., Liu, X., Meyerowitz, J.T., Snow, C.D., Chen, M.M. and Arnold, F.H.
554 Engineered ketol-acid reductoisomerase and alcohol dehydrogenase enable anaerobic 2-

methylpropan-1-ol production at theoretical yield in *Escherichia coli*. *Metabolic Engineering* **13**, 345-352 (2011).
41. Ghosh, I.N., Martien, J., Hebert, A.S., Zhang, Y., Coon, J.J., Amador-Noguez, D. and Landick, R. OptSSeq explores enzyme expression and function landscapes to maximize isobutanol production rate. *Metabolic Engineering* **52**, 324-340 (2019).
42. Akita, H., Nakashima, N. & Hoshino, T. Bacterial production of isobutanol without expensive reagents. *Applied Microbiology and Biotechnology* **99**, 991-999 (2015).
43. Baret, J.-C. et al. Fluorescence-activated droplet sorting (FADS): efficient microfluidic cell sorting based on enzymatic activity. *Lab on a Chip* **9**, 1850-1858 (2009).
44. Clowes, R.C. Molecular structure of bacterial plasmids. *Bacteriological Reviews* **36**, 361 (1972).
45. Ebersbach, G. & Gerdes, K. Plasmid segregation mechanisms. *Annual Review of Genetics* **39**, 453-479 (2005).
46. Kachroo, A.H., Jayaram, M. & Rowley, P.A. Metabolic engineering without plasmids. *Nature Biotechnology* **27**, 729 (2009).
47. Keasling, J.D. Gene-expression tools for the metabolic engineering of bacteria. *Trends in Biotechnology* **17**, 452-460 (1999).
48. Arfin, S.M., Ratzkin, B. & Umbarger, H. The metabolism of valine and isoleucine in *Escherichia coli.* XVII. The role of induction in the depression of acetohydroxy acid isomeroreductase. *Biochemical and Biophysical Research Communications* **37**, 902-908 (1969).
49. Wang, T. & Dunlop, M.J. Controlling and exploiting cell-to-cell variation in metabolic engineering. *Current Opinion in Biotechnology* **57**, 10-16 (2019).
50. Atsumi, S. et al. Engineering the isobutanol biosynthetic pathway in *Escherichia coli* by comparison of three aldehyde reductase/alcohol dehydrogenase genes. *Applied Microbiology and Biotechnology* **85**, 651-657 (2010).
51. Livak, K.J. & Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods* **25**, 402-408 (2001).
52. Smith, K.M. & Liao, J.C. An evolutionary strategy for isobutanol production strain development in Escherichia coli. *Metabolic Engineering* **13**, 674-681 (2011).
53. Elowitz, M.B., Levine, A.J., Siggia, E.D. & Swain, P.S. Stochastic gene expression in a single cell. *Science* **297**, 1183-1186 (2002).
54. Zhou, K. et al. Novel reference genes for quantifying transcriptional responses of *Escherichia coli* to protein overexpression by quantitative PCR. *BMC Molecular Biology* **12**, 18-18 (2011).

Acknowledgments

General: The authors thank Scott Scholz and Peter Freddolino for valuable discussions regarding position-dependent gene expression variation and transposon footprinting, as well as Mark Burns, Lola Eniola-Adefeso, and Chris Barr for the use of equipment in their laboratories. The isobutanol strains and plasmids were originally provided by James Liao (UCLA, formerly). The CIChE construct was constructed using the pTGD plasmid provided by Keith Tyo (Northwestern U).

Funding: This work was supported by the USDA AFRI NIFA Fellowships Grant Program (Grant no. 2016-67011-24725).

Author contributions: T.S. and X.N.L. conceived and designed the study. T.S. and A.K. constructed the integration libraries. M.T.C. and K.K. developed the droplet sorting platform. T.S. and M.T.C. performed the library screening. T.S. and D.N.C. characterized the library isolates and fully integrated strains. T.S. and X.N.L. wrote the paper.

Competing interests: The authors declare no competing interests.

Figures and Tables

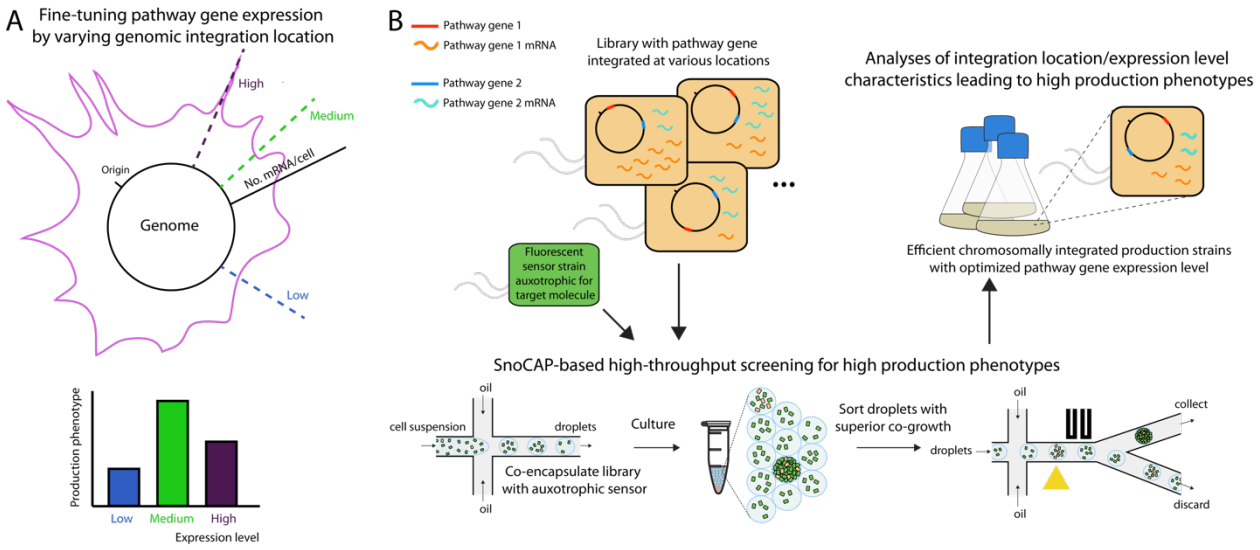


Figure 1. New approach for optimizing chromosomal heterologous gene expression by position-dependent expression variation. (a) Gene expression level varies depending on distance from the origin of replication, as well as other position-dependent factors. The expression level of a heterologous pathway gene affects the production phenotype of a strain. Increasing the expression level of a gene can improve production level by increasing flux through the pathway, but avoiding too high of an expression level can also provide benefits such as reducing the buildup of toxic intermediates and avoiding unnecessary metabolic burden. (b) Libraries with pathway genes integrated into various genome locations can be screened for production phenotype using syntrophic co-culture amplification of production (SnoCAP)[36]. A cross-feeding co-culture is configured consisting of a secretor strain producing a target molecule and a sensor strain that cannot synthesize the target molecule autonomously. The co-cultures are compartmentalized in microfluidic droplets, incubated to allow co-growth, and then sorted to identify the strains with high production phenotypes.

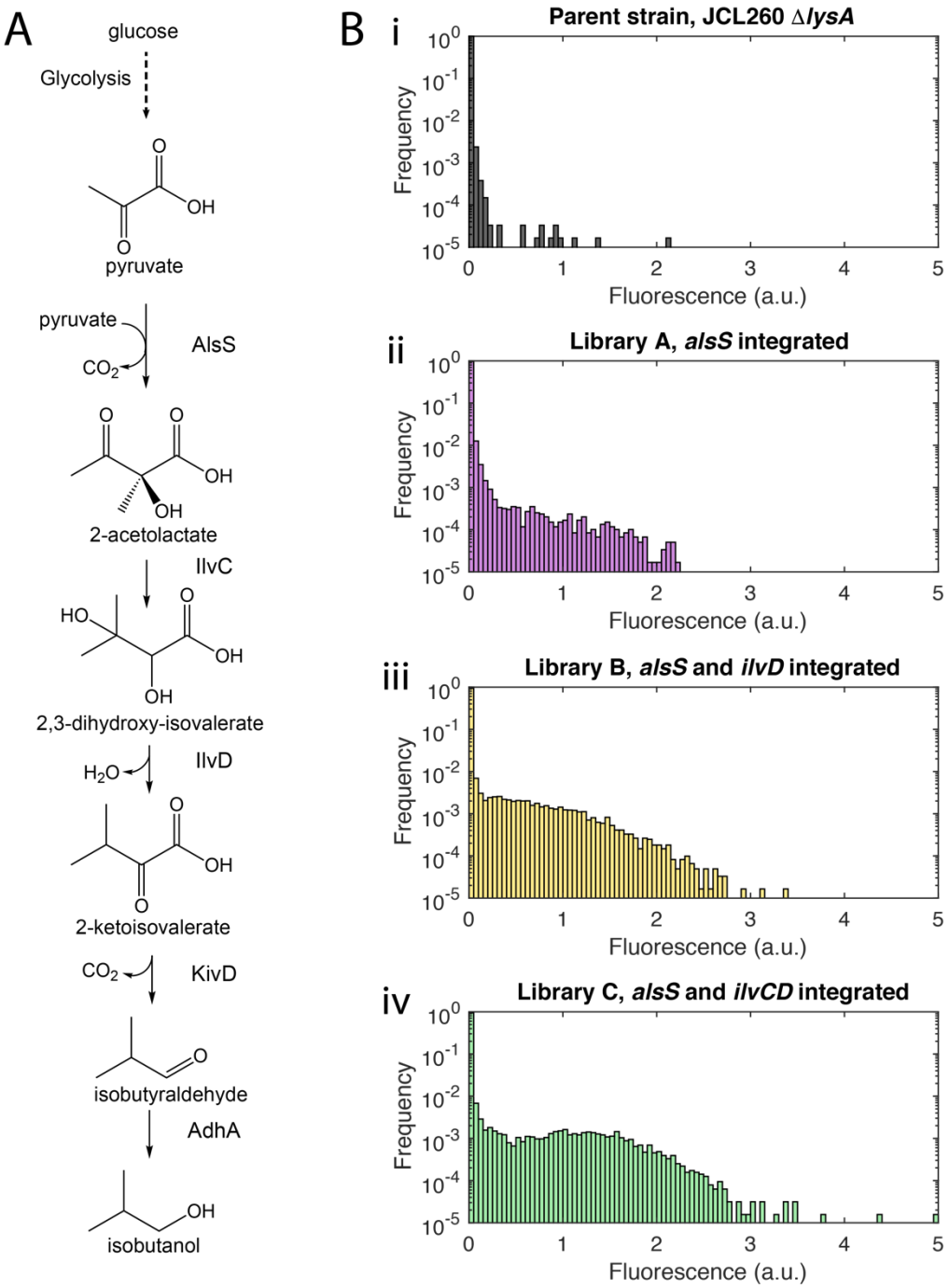


Figure 2. Screening of libraries with *alsS*, *ilvCD* integrated by Tn5. (a) Isobutanol production pathway[50]. Acetolactate synthase (AlsS), ketol-acid reductoisomerase (IlvC), dihydroxy-acid dehydratase (IlvD), 2-ketoacid decarboxylase (KivD), and alcohol dehydrogenase (AdhA). (b) Histograms comparing fluorescence signal from droplets containing co-cultures of fluorescent sensor strain with unlabeled secretor strain libraries, compared to the parent strain (i). The libraries were constructed by Tn5 integration of *kan*-PLlacO1-*alsS* (Library A) (ii), PLlacO1-*ilvD* followed by *kan*-PLlacO1-*alsS* (Library B) (iii), or PLlacO1-*ilvCD* followed by *kan*-PLlacO1-*alsS* (Library C) (iv). Each histogram comprises data from the analysis of about 60,000 droplets.

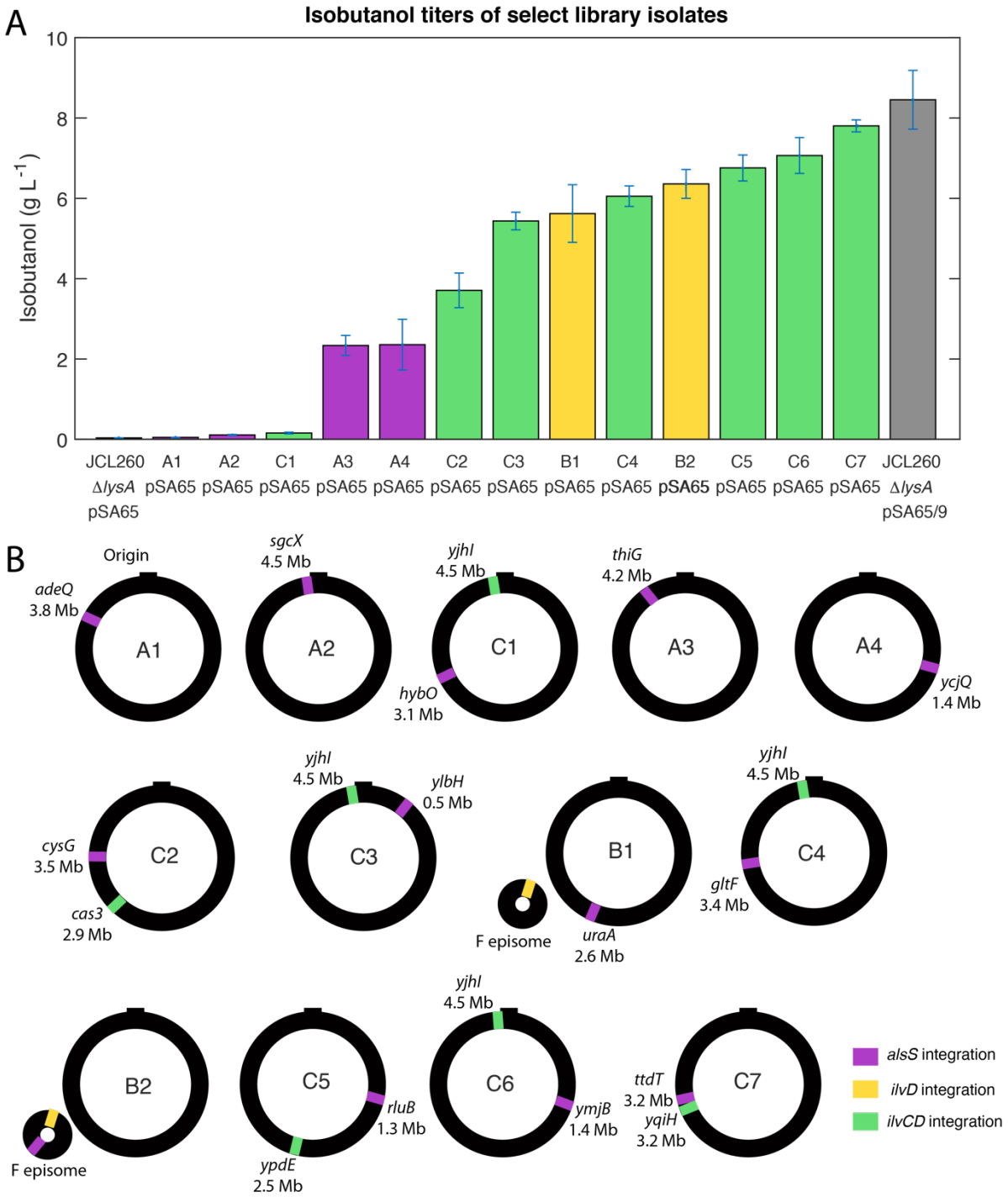


Figure 3. Production phenotype and gene insertion locations of select library isolates. a) Isobutanol titers of library isolates displaying various production levels. Samples were taken 48 h after induction of the pathway genes. Each of the isolates has been transformed with the pSA65 plasmid expressing *kivD* and *adhA*. Parent strain JCL260 Δ*lysA* and JCL260 Δ*lysA* pSA69, which expresses the *alsS* and *ilvCD* genes from the pSA69 plasmid, each also carrying pSA65, are included for comparison. Error bars represent the standard deviation of three biological replicates. (b) Gene integration locations of the isolates whose production is shown in (a).

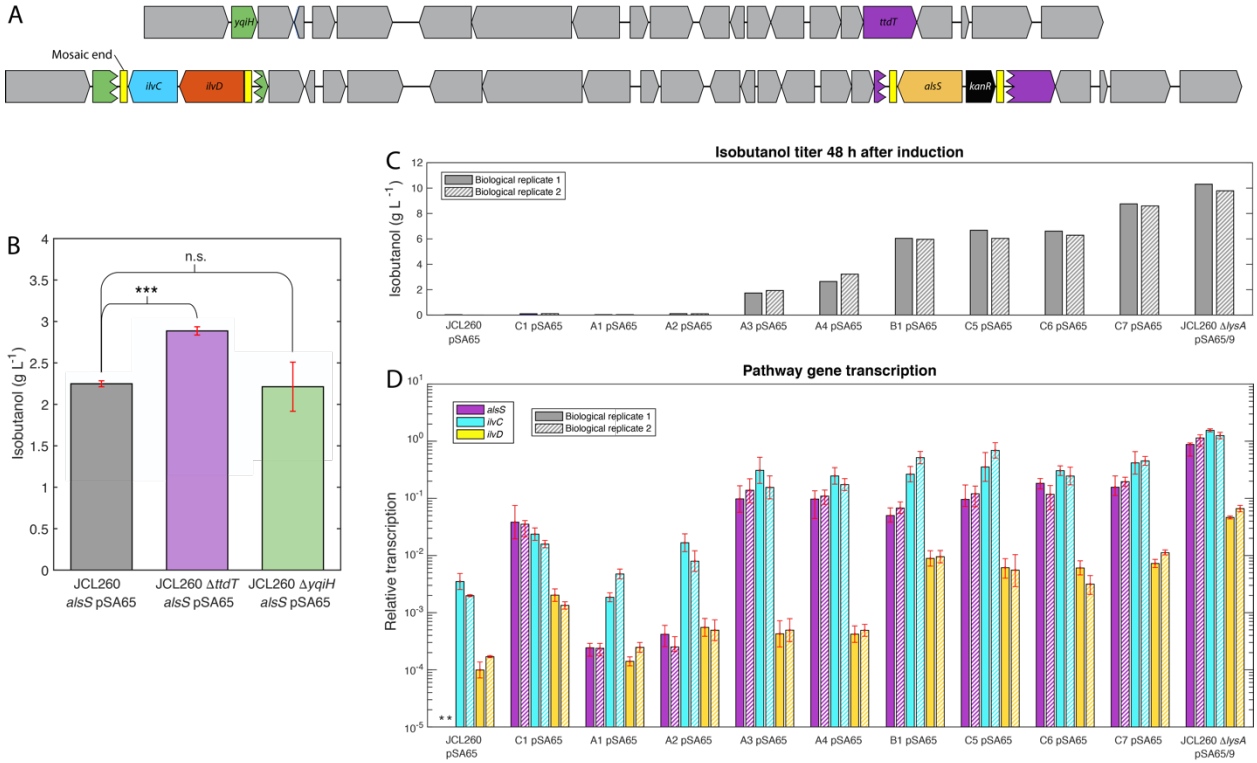


Figure 4. Investigation of mechanisms underlying production phenotypes of select library isolates. (A) Schematic of the locations of *kan*-PLlacO1-*alsS* and PLlacO1-*ilvCD* integrations in isolate C7, determined by transposon footprinting. (B) Isobutanol production of strains with *alsS* genomically integrated, containing plasmid pSA65 encoding *kivD* and *adhA*, and possessing deletions of either *ttdT* or *yqiH*, compared to the same strain without either deletion. Data shown are from samples taken 72 h after induction with IPTG. ***: significant difference at *P*-value < 0.0001 using an unpaired t-test. n.s.: not significant. (C) Isobutanol production 48 h after induction of the cultures from which the RNA had been harvested. (D) Expression levels of *alsS*, *ilvC*, and *ilvD*, relative to *alsS* level in JCL260 Δ*lysA* pSA65/9. RNA was harvested 3 h after induction with IPTG. * indicates no amplification during the qPCR reaction. Data from two biological replicate cultures of each strain are shown. Error bars in (B) represent the standard deviation of three biological replicates. Error bars in (D) represent the standard deviation of three technical replicates.

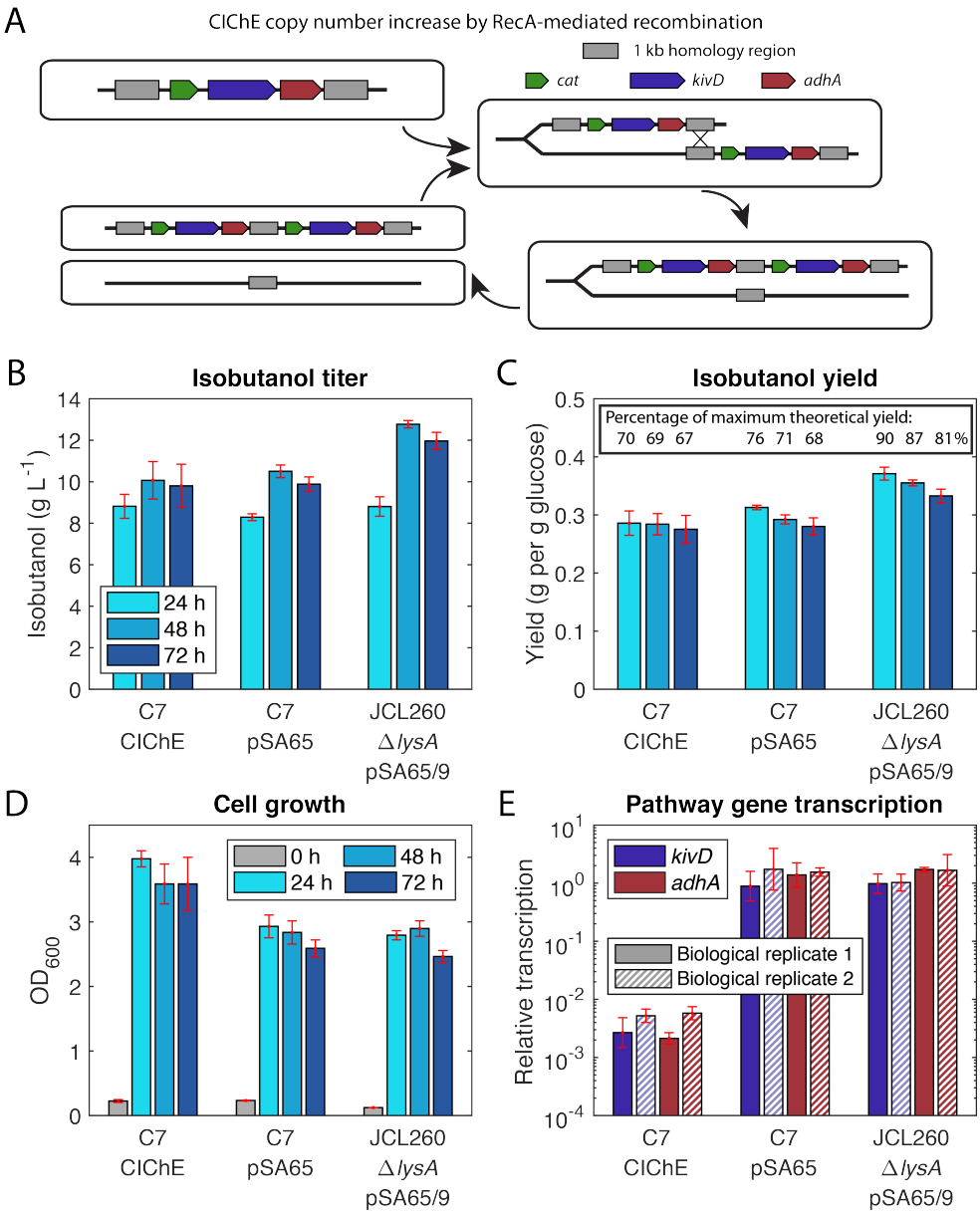


Figure 5. Fully integrated strain performance compared to strains with plasmids. (a) Overview of the chemically inducible chromosomal evolution (CIChE) approach for gene copy number amplification by RecA-mediated homologous recombination. (b) Isobutanol production by strain C7 after integration of *kivD-adhA* by CIChE, compared to C7 pSA65, which has *alsS* and *ilvCD* integrated into the genome and *kivD* and *adhA* expressed from the pSA65 plasmid, and JCL260 Δ*lysA* pSA65/9 which has all five pathway genes expressed from the plasmids. Error bars represent the standard deviation of three biological replicates. (c) The yields of the cultures based on glucose consumed. The inset in (c) lists the percentage of the theoretical maximum yield (0.41 g isobutanol per gram glucose). (d) The cell densities of the production cultures whose performance is displayed in (b) and (c). (e) Expression levels of *kivD* and *adhA*, relative to *kivD* level in JCL260 Δ*lysA* pSA65/9. RNA was harvested 4 h after induction with IPTG. Error bars represent the standard deviation of three technical replicates.

692 **Table 1.** Gene integration locations.
693

Strain	Integration construct	Gene	Coordinate in gene of integration location	Direction relative to gene integrated into
A1	*kan-alsS*	*adeQ*	436	Opposite
A2	*kan-alsS*	*sgcX*	111	Opposite
C1	*kan-alsS*	*hybO*	812	Opposite
C1	*ilvCD*	*yjhI*	231	Same
A3	*kan-alsS*	*thiG*	73	Same
A4	*kan-alsS*	*ycjQ*	364	Opposite
C2	*kan-alsS*	*cysG*	563	Opposite
C2	*ilvCD*	*cas3*	1,387	Same
C3	*kan-alsS*	*ylbH*	205	Opposite
C3	*ilvCD*	*yjhI*	231	Same
B1	*kan-alsS*	*uraA*	659	Same
B1	*ilvD*	gamma delta (Tn1000) region of F plasmid	426 bp into *tnpX*	Same as *tnpX*
C4	*kan-alsS*	*gltF*	313	Opposite
C4	*ilvCD*	*yjhI*	231	Same
B2	*kan-alsS*	Upstream of *psiB* on F plasmid	218 bp upstream	Opposite of *psiB*
B2	*ilvD*	gamma delta (Tn1000) region of F plasmid	426 bp into *tnpX*	Same as *tnpX*
C5	*kan-alsS*	*rluB*	794	Same
C5	*ilvCD*	*ypdE*	76	Same
C6	*kan-alsS*	*ymjB*	7	Opposite
C6	*ilvCD*	*yjhI*	231	Same
C7	*kan-alsS*	*ttdT*	114	Opposite
C7	*ilvCD*	*yqiH*	512	Opposite

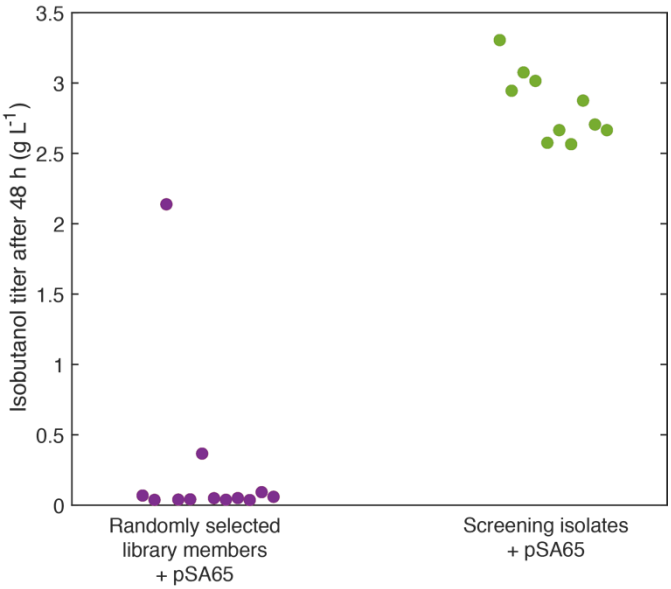
694
695
696

697 **Supplementary Materials**

698

699 **Supplementary Figures**

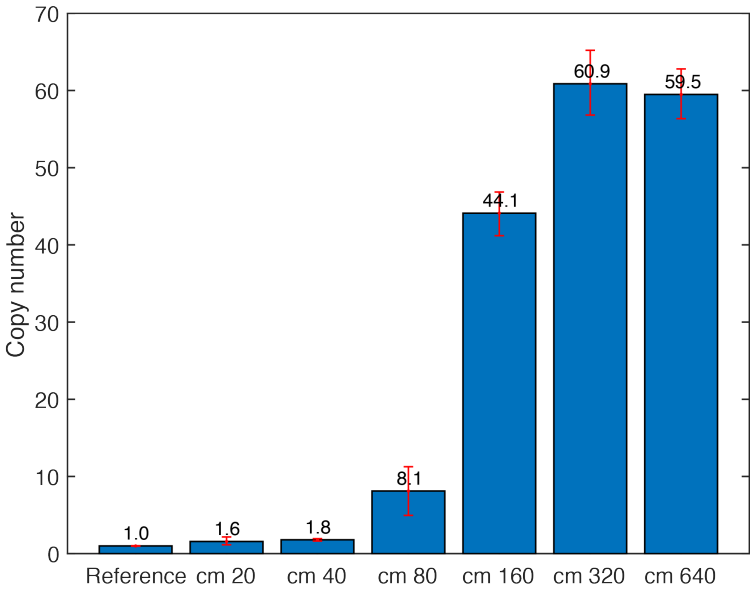
700



701

702 **Figure S1.** Comparison of production performance of randomly selected library members to
703 library members isolated by screening, after transformation with pSA65. Plot shows isobutanol
704 titers 48 h after IPTG induction of 12 randomly selected colonies from unscreened Library A (*alsS*
705 integrated) and 10 colonies from the pool of droplets produced by screening Library A, after
706 sorting the top 0.3% most fluorescent droplets (top 3% of library members).

707



708

709 **Figure S2.** CIChE construct copy number. Copy number of a CIChE construct containing the
710 *kivD, adhA,* and *cat* genes in the NV3r1 strain after passaging into higher concentrations of
711 chloramphenicol and deletion of *recA*. The measurement was conducted by qPCR of the *cat* gene.
712 "cm X" refers to NV3r1 with the CIChE construct integrated in the *aslB* site and propagated in X
713 μg/mL of chloramphenicol before deletion of the *recA* gene. "Reference" refers to a strain with a
714 single copy of the *cat* gene (NV3r1 Δ*intC*::*yfp-cat*). Error bars represent the standard deviation of
715 three technical replicates.

716 **Supplementary Tables**

717
718 **Table S1: Strains and plasmids**
719

Strain	Relevant genotype	Reference
JCL260	BW25113/F′ [*traD36*, *proAB*+, *lacIq* ZΔM15 (TetR)] Δ*ldhA* Δ*frd* Δ*adhE* Δ*pta* Δ*fnr* Δ*pflB*	[38]
JCL260 Δ*lysA*	JCL260 Δ*lysA::kan*	[36]
JCL260 Δ*lysA alsS*	JCL260 Δ*lysA::*FRT with $P_L$lacO_1::*alsS* integrated between *yghX* and *gpr*	[36]
JCL260 Δ*lysA* Δ*ttdT alsS*	JCL260 Δ*lysA::*FRT Δ*ttdT::*FRT with $P_L$lacO_1::*alsS* integrated between *yghX* and *gpr*	This study
JCL260 Δ*lysA* Δ*yqiH alsS*	JCL260 Δ*lysA::*FRT Δ*yqiH::*FRT with $P_L$lacO_1::*alsS* integrated between *yghX* and *gpr*	This study
JCL260 Δ*lysA* pSA65/9	JCL260 Δ*lysA::*FRT pSA65/9	[36]
JCL260 Δ*lysA* Δ*ilvD*	JCL260 Δ*lysA::*FRT Δ*ilvD::*FRT	This study
C7	JCL260 Δ*lysA::*FRT Δ*ilvD::*FRT Δ*ttdT::kan-alsS* Δ*yqiH::ilvCD*	This study
JCL260 CIChE	JCL260 Δ*lysA::*FRT with *adhA-kivD* CIChE construct integrated in the *aslB* site	[36]
C7 CIChE	C7 with *adhA-kivD* CIChE construct integrated in the *aslB* site	This study
K12 Δ*ilvD*	K12 Δ*ilvD::kan*	[36]
K12 Δ*ilvD cfp* pSAS31	K12 Δ*ilvD::*FRT Δ*galK::cfp-bla* pSAS31	[36]
Δ*ilvD* Keio strain	BW25113 Δ*ilvD::kan* (JW5605-1)	Coli Genetic Stock Center (CGSC)
Δ*ttdT* Keio strain	BW25113 Δ*ttdT::kan* (JW3035-1)	CGSC
Δ*yqiH* Keio strain	BW25113 Δ*yqiH::kan* (JW5508-3)	CGSC
BW 26,547	Δ*recA*::*kan* Lambda *recA*+	CGSC
Reference	NV3r1 Δ*intC*::*yfp-cat*	NV3r1[52] Δ*intC*::*yfp-cat*[53], transferred into NV3r1 by P1 transduction
cm X	NV3r1 with *adhA-kivD* CIChE construct integrated in the *aslB* site, maintained	[52]

	both on plates and in liquid medium with X µg/mL chloramphenicol	
Plasmid		
pSA65	ColE1 ori; AmpR; $P_L lacO_1$::*kivd–adhA*	[50]
pSA69	p15A ori; KanR; $P_L lacO_1$::*alsS–ilvC–ilvD*	[38]
pSAS31	Constitutively expressed mNeonGreen plasmid; kanamycin-resistant	[33]

720
721
722 **Table S2: Primers and oligonucleotides**
723

Name	Sequence	Notes
	Tn5 integration construct primers	
phosph_transp_kan_for	**5Phos/**CTGTCTCTTATACACATCTACCGAGCGTTCTGAACAAAT	**5Phos/** indicates 5' phosphorylation Underlined region indicates the added Tn5 mosaic end Blue highlighting indicates added LlacO1 promoter Yellow highlighting indicates homology to the pSA69 plasmid
phosph_transp_ilvD_rev	**5Phos/**CTGTCTCTTATACACATCTGCGGATTTGTCCTACTCAGG	
phosph_transp_alsS_rev	**5Phos/**CTGTCTCTTATACACATCTGGTGATTCCTCGTCGACCTA	
phosph_transp_pLlacO1_ilvC_for	**5Phos/**CTGTCTCTTATACACATCTAATTGTGAGCGGATAACAATTGACATTGTGAGCGGATAACAAGATACTGAGCACATCAGCAGGACGCACTGACCGTCGACGAGGAATCACCAT	
phosph_transp_pLlacO1_ilvD_for:	**5Phos/**CTGTCTCTTATACACATCTAATTGTGAGCGGATAACAATTGACATTGTGAGCGGATAACAAGATACTGAGCACATCAGCAGGACGCACTGACCCCCGGAAGGAGATATACC	
	Y-linker oligos	
phos_CviQI_YTA3	**5Phos/**TAGACGCTACGTCCGTGTTGTCGGTCCTG	CviQI Y-Linker primer **5Phos/** indicates 5' phosphorylation

CviQI_YTA5	ACTACGCACGCGACGAGACGTAGCGTC	CviQI Y-Linker primer for annealing with phos_CviQI_YTA3
phos_CviAII_YAT3	**5Phos/**ATCACGCTACGTCCGTGTTGTCGGTCCTG	CviAII Y-Linker primer **5Phos/** indicates 5' phosphorylation
CviAII_YAT5	ACTACGCACGCGACGAGACGTAGCGTG	CviAII Y-Linker primer for annealing with phos_CviAII_YAT3
Footprinting primers		
Y_linker_primer	ACTACGCACGCGACGA	
kan_alsS_front_footprint_rev	TCGCCTTCTTGACGAGTTCT	
PLlacO1_footprint_rev	TCCGCTCACAATTAGATGTGTAT	
Gene deletion verification primers		
ilvD_verif_for	GTACCTGGCGGATGAAGTGT	
ilvD_verif_rev	TTATCAAGACGCGACGACAG	
ttdT_verif_for	CCCAGGTGGAAGAGATTGAA	
ttdT_verif_rev	GAGTTACCGGCTGCGTAAAA	
yqiH_verif_for	GCCAAAATTAATTCAGGATGAAA	
yqiH_verif_rev	GCCCATGCAGGTAATGAACTA	
qPCR Primers		
alsS_qPCR_for	GATGCGGTATTTGACGCTTT	123 bp product

alsS_qPCR_rev	TAACACGACTCCCGGTTTTC	
ilvC_qPCR_for	AGGTACCGATCCAGCATACG	104 bp product
ilvC_qPCR_rev	AGAGACGGTCCATCATCAGG	
ilvD_qPCR_for	TATCGGTATTCTCGGCGAAC	114 bp product
ilvD_qPCR_rev	TGGGTCAGCATAACGTCGTA	
cysG_qPCR_for	TTGTCGGCGGTGGTGATGTC	105 bp product From [54]
cysG_qPCR_rev	ATGCGGTGAACTGTGGAATAAACG	
Integration verification primers		
alsS_int_chk_front_for	ACCTCTCCTTTCCACCGTTC	Located in *gpr* gene
alsS_int_chk_front_rev	TCGCCTTCTTGACGAGTTCT	Located in *kan* gene
alsS_int_chk_back_for	GGGGAACTCATGAAAACGAA	Located in *alsS* gene
alsS_int_chk_back_rev	GAGATTTTCCCGTGAGCGTA	Located in *yghX* gene

724
725

Supplementary Information for

Optimized gene expression from bacterial chromosome by high-throughput integration and screening

Tatyana E. Saleski[1], Meng Ting Chung[2], David N. Carruthers[1], Azzaya Khasbaatar[1], Katsuo Kurabayashi[2,3], Xiaoxia Nina Lin[1,4]*

[1]Department of Chemical Engineering, University of Michigan, Ann Arbor, MI, USA
[2]Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI, USA
[3]Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA
[4]Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI, USA

*Corresponding author. Email: ninalin@umich.edu

This PDF file includes:

Fig. S1 and S2……1
Tables S1 and S2……2
References……3

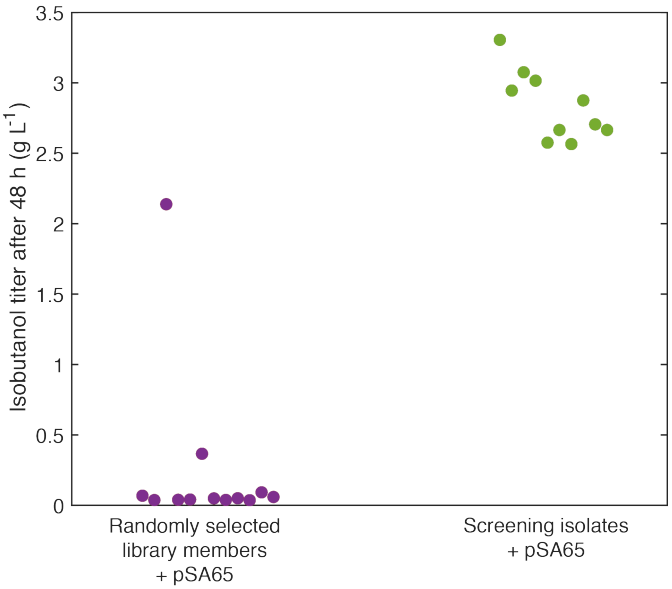


Figure S1. Comparison of production performance of randomly selected library members to library members isolated by screening, after transformation with pSA65. Plot shows isobutanol titers 48 h after IPTG induction of 12 randomly selected colonies from unscreened Library A (*alsS* integrated) and 10 colonies from the pool of droplets produced by screening Library A, after sorting the top 0.3% most fluorescent droplets (top 3% of library members).

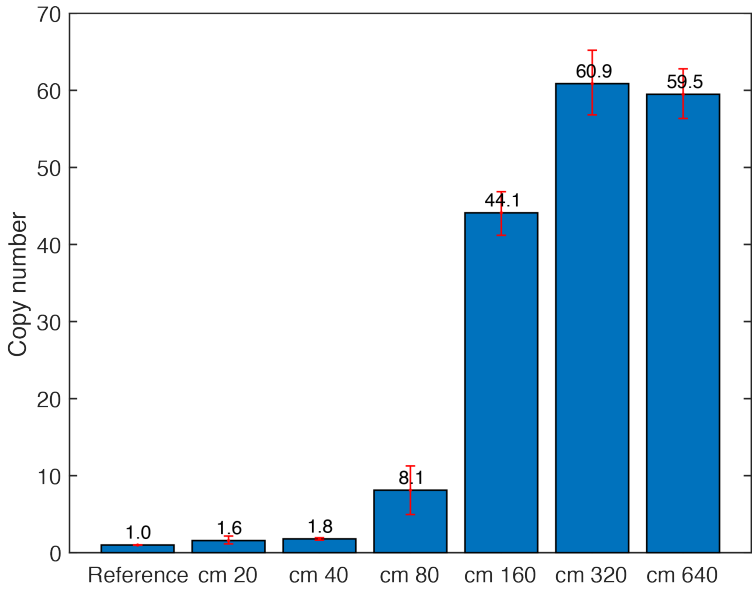


Figure S2. CIChE construct copy number. Copy number of a CIChE construct containing the *kivD, adhA,* and *cat* genes in the NV3r1 strain after passaging into higher concentrations of chloramphenicol and deletion of *recA*. The measurement was conducted by qPCR of the *cat* gene. "cm X" refers to NV3r1 with the CIChE construct integrated in the *aslB* site and propagated in X µg/mL of chloramphenicol before deletion of the *recA* gene. "Reference" refers to a strain with a single copy of the *cat* gene (NV3r1 Δ*intC*::*yfp-cat*). Error bars represent the standard deviation of three technical replicates.

Supplementary Tables

Table S1: Strains and plasmids

Strain	Relevant genotype	Reference
JCL260	BW25113/F′ [*traD36*, *proAB*+, *lacIq* ZΔM15 (TetR)] Δ*ldhA* Δ*frd* Δ*adhE* Δ*pta* Δ*fnr* Δ*pflB*	[1]
JCL260 Δ*lysA*	JCL260 Δ*lysA::kan*	[2]
JCL260 Δ*lysA alsS*	JCL260 Δ*lysA::*FRT with $P_L$lacO_1::*alsS* integrated between *yghX* and *gpr*	[2]
JCL260 Δ*lysA* Δ*ttdT alsS*	JCL260 Δ*lysA::*FRT Δ*ttdT::*FRT with $P_L$lacO_1::*alsS* integrated between *yghX* and *gpr*	This study
JCL260 Δ*lysA* Δ*yqiH alsS*	JCL260 Δ*lysA::*FRT Δ*yqiH::*FRT with $P_L$lacO_1::*alsS* integrated between *yghX* and *gpr*	This study
JCL260 Δ*lysA* pSA65/9	JCL260 Δ*lysA::*FRT pSA65/9	[2]
JCL260 Δ*lysA* Δ*ilvD*	JCL260 Δ*lysA::*FRT Δ*ilvD::*FRT	This study
C7	JCL260 Δ*lysA::*FRT Δ*ilvD::*FRT Δ*ttdT::kan-alsS* Δ*yqiH::ilvCD*	This study
JCL260 CIChE	JCL260 Δ*lysA::*FRT with *adhA-kivD* CIChE construct integrated in the *aslB* site	[2]
C7 CIChE	C7 with *adhA-kivD* CIChE construct integrated in the *aslB* site	This study
K12 Δ*ilvD*	K12 Δ*ilvD::kan*	[2]
K12 Δ*ilvD cfp* pSAS31	K12 Δ*ilvD::*FRT Δ*galK::cfp-bla* pSAS31	[2]
Δ*ilvD* Keio strain	BW25113 Δ*ilvD::kan* (JW5605-1)	Coli Genetic Stock Center (CGSC)
Δ*ttdT* Keio strain	BW25113 Δ*ttdT::kan* (JW3035-1)	CGSC
Δ*yqiH* Keio strain	BW25113 Δ*yqiH::kan* (JW5508-3)	CGSC
BW 26,547	Δ*recA*::*kan* Lambda *recA*+	CGSC
Reference	NV3r1 Δ*intC*::*yfp-cat*	NV3r1[3]

		Δ*intC*::*yfp-cat*[4], transferred into NV3r1 by P1 transduction
cm X	NV3r1 with *adhA-kivD* CIChE construct integrated in the *aslB* site, maintained both on plates and in liquid medium with X µg/mL chloramphenicol	[3]
Plasmid		
pSA65	ColE1 ori; AmpR; $P_L lacO_1$::*kivd–adhA*	[5]
pSA69	p15A ori; KanR; $P_L lacO_1$::*alsS–ilvC–ilvD*	[1]
pSAS31	Constitutively expressed mNeonGreen plasmid; kanamycin-resistant	[6]

Table S2: Primers and oligonucleotides

Name	Sequence	Notes
Tn5 integration construct primers		
phosph_transp_kan_for	**5Phos/**CTGTCTCTTATACACATCTACCGAGCGTTCTGAACAAAT	**5Phos/** indicates 5' phosphorylation Underlined region indicates the added Tn5 mosaic end Blue highlighting indicates added LlacO1 promoter Yellow highlighting indicates homology to the pSA69 plasmid
phosph_transp_ilvD_rev	**5Phos/**CTGTCTCTTATACACATCTGCGGATTTGTCCTACTCAGG	
phosph_transp_alsS_rev	**5Phos/**CTGTCTCTTATACACATCTGGTGATTCCTCGTCGACCTA	
phosph_transp_pLlacO1_ilvC_for	**5Phos/**CTGTCTCTTATACACATCTAATTGTGAGCGGATAACAATTGACATTGTGAGCGGATAACAAGATACTGAGCACATCAGCAGGACGCACTGACCGTCGACGAGGAATCACCAT	
phosph_transp_pLlacO1_ilvD_for:	**5Phos/**CTGTCTCTTATACACATCTAATTGTGAGCGGATAACAATTGACATTGTGAGCGGATAACAAGATACTGAGCACATCAGCAGGACGCACTGACCCCCGGAAGGAGATATACC	

Y-linker oligos		
phos_CviQI_YTA3	**5Phos/**TAGACGCTACGTCCGTGTTGTCGGTCCTG	CviQI Y-Linker primer **5Phos/** indicates 5' phosphorylation
CviQI_YTA5	ACTACGCACGCGACGAGACGTAGCGTC	CviQI Y-Linker primer for annealing with phos_CviQI_YTA3
phos_CviAII_YAT3	**5Phos/**ATCACGCTACGTCCGTGTTGTCGGTCCTG	CviAII Y-Linker primer **5Phos/** indicates 5' phosphorylation
CviAII_YAT5	ACTACGCACGCGACGAGACGTAGCGTG	CviAII Y-Linker primer for annealing with phos_CviAII_YAT3
Footprinting primers		
Y_linker_primer	ACTACGCACGCGACGA	
kan_alsS_front_footprint_rev	TCGCCTTCTTGACGAGTTCT	
PLlacO1_footprint_rev	TCCGCTCACAATTAGATGTGTAT	
Gene deletion verification primers		
ilvD_verif_for	GTACCTGGCGGATGAAGTGT	
ilvD_verif_rev	TTATCAAGACGCGACGACAG	
ttdT_verif_for	CCCAGGTGGAAGAGATTGAA	
ttdT_verif_rev	GAGTTACCGGCTGCGTAAAA	

yqiH_verif_for	GCCAAAATTAATTCAGGATGAAA	
yqiH_verif_rev	GCCCATGCAGGTAATGAACTA	
qPCR Primers		
alsS_qPCR_for	GATGCGGTATTTGACGCTTT	123 bp product
alsS_qPCR_rev	TAACACGACTCCCGGTTTTC	
ilvC_qPCR_for	AGGTACCGATCCAGCATACG	104 bp product
ilvC_qPCR_rev	AGAGACGGTCCATCATCAGG	
ilvD_qPCR_for	TATCGGTATTCTCGGCGAAC	114 bp product
ilvD_qPCR_rev	TGGGTCAGCATAACGTCGTA	
cysG_qPCR_for	TTGTCGGCGGTGGTGATGTC	105 bp product From [7]
cysG_qPCR_rev	ATGCGGTGAACTGTGGAATAAACG	
Integration verification primers		
alsS_int_chk_front_for	ACCTCTCCTTTCCACCGTTC	Located in *gpr* gene
alsS_int_chk_front_rev	TCGCCTTCTTGACGAGTTCT	Located in *kan* gene
alsS_int_chk_back_for	GGGGAACTCATGAAAACGAA	Located in *alsS* gene
alsS_int_chk_back_rev	GAGATTTTCCCGTGAGCGTA	Located in *yghX* gene

References

1. Atsumi, S., Hanai, T. & Liao, J.C. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **451**, 86 (2008).
2. Saleski, T.E. et al. Syntrophic co-culture amplification of production phenotype for high-throughput screening of microbial strain libraries. *Metabolic Engineering* **54**, 232-243 (2019).
3. Smith, K.M. & Liao, J.C. An evolutionary strategy for isobutanol production strain development in Escherichia coli. *Metabolic Engineering* **13**, 674-681 (2011).
4. Elowitz, M.B., Levine, A.J., Siggia, E.D. & Swain, P.S. Stochastic gene expression in a single cell. *Science* **297**, 1183-1186 (2002).
5. Atsumi, S. et al. Engineering the isobutanol biosynthetic pathway in *Escherichia coli* by comparison of three aldehyde reductase/alcohol dehydrogenase genes. *Applied Microbiology and Biotechnology* **85**, 651-657 (2010).
6. Scholz, S.A. et al. High-resolution mapping of the *Escherichia coli* chromosome reveals positions of high and low transcription. *Cell Systems* **8**, 212-225 (2019).
7. Zhou, K. et al. Novel reference genes for quantifying transcriptional responses of *Escherichia coli* to protein overexpression by quantitative PCR. *BMC Molecular Biology* **12**, 18-18 (2011).